UNIVERSITY OF ALBERTA
LIBRARY

Regulations Regarding Theses and Dissertations

Typescript copies of theses and dissertations for Master's and Doctor's degrees deposited in the University of Alberta Library, as the official Copy of the Faculty of Graduate Studies, may be consulted in the Reference Reading Room only.

A second copy is on deposit in the Department under whose supervision the work was done. Some Departments are willing to loan their copy to libraries, through the inter-library loan service of the University of Alberta Library.

These theses and dissertations are to be used only with due regard to the rights of the author. Written permission of the author and of the Department must be obtained through the University of Alberta Library when extended passages are copied. When permission has been granted, acknowledgement must appear in the published work.

This thesis or dissertation has been used in accordance with the above regulations by the persons listed below. The borrowing library is obligated to secure the signature of each user.

Please sign below:

Date	Signature	Institution

THE UNIVERSITY OF ALBERTA

A RATING SCALE FOR DISCRIMINATING RELATIVE PLAYING PERFORMANCE OF SKILLED FEMALE VOLLEYBALL PLAYERS

by

PATRICIA L. JACKSON

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE

REQUIREMENTS FOR THE DEGREE

OF MASTER OF ARTS

FACULTY OF PHYSICAL EDUCATION

EDMONTON, ALBERTA

OCTOBER, 1967

THE UNIVERSITY OF

by

A THESIS

OF MASTER OF ARTS

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "A Rating Scale for Discriminating Relative Playing Performance of Skilled Female Volleyball Players," submitted by Patricia L. Jackson in partial fulfilment of the requirements for the degree of Master of Arts.

Date ..... Nov. 1, 1967 ...

ABSTRACT

It was the purpose of this study to develop a valid, objective and practical measuring instrument that would discriminate relative volleyball performance of skilled female players in a competitive game situation.

Data was collected from two separate volleyball tournaments. A total of one hundred sixteen individuals were rated during seventeen matches at the Second Century Week Tournament and the Canadian Senior Women's Volleyball Championships.

Face validity of the instrument was shown by demonstrating that the items included in the Volleyball Rating Scale were important to successful performance in skilled volleyball competitions. Statistical validity did not substantiate the demonstrated curricular validity when results of the Volleyball Rating Scale were correlated with the averaged rankings of a panel of judges. The reported validity coefficients were .109 and .470. In the first instance no consideration was given to the number of games played by each player in a match. The second validity coefficient was determined after consideration was given to the number of games played by each player per match.

The use of multiple regression techniques indicated that the two separate panels of judges used in this study did not consider the items on the Volleyball Rating Scale to be of equal importance. One panel weighted the items; good serve 4, ace 3, pass 2, set 1, spike 1,

block 1, return -2, and poor serve -2. The second panel weighted the items; good serve 8, pass 7, spike 3, ace 2, return 1, poor serve -1, block -4, and set -4.

It was found that set, pass, return, block, violations, ace and poor serve discriminated between good and poor performances as determined by the Flanagan technique. Good serve and spike did not significantly discriminate performance.

The reliability of the Volleyball Rating Scale, determined by correlating the rankings of game one with those of game two was found to be .395 ($p<.05$). An averaged objectivity coefficient of .876 was reported, significant beyond the .01 probability level.

Within the limitations of the statistical procedures employed, the experimental design, the samples investigated, and the personnel serving as judges, the following conclusions were made.

1. The Volleyball Rating Scale possessed curricular validity. Statistical validity was not significant at the .05 level.

2. The objectivity of the Volleyball Rating Scale was significant at the .01 level.

3. Evaluation of relative performance by means of the Volleyball Rating Scale was found to be practical in that only two persons were required to obtain objective results.

4. The Volleyball Rating Scale can be used either separately or in conjunction with other evaluation methods for determination of

relative performance of skilled female volleyball players.

5. The Volleyball Rating Scale may be used for determination of individual or team strengths and weaknesses with respect to the items recorded.

6. The Volleyball Rating Scale has two major advantages over conventional methods of volleyball evaluation. It measures a realistic competitive situation and it is diagnostic.

7. No significant differences were found between the Clifton Single Hit Volley Test and the Volleyball Rating Scale as methods of determining relative game performance of skilled female players.

The results of the study suggested the following recommenations:

1. That further investigations be made with the instrument using panels of judges consisting of at least five members to determine if statistical validity can be improved.

2. That further study be undertaken to determine if the relative importance and weighting of the Volleyball Rating Scale items can be accurately established.

3. That similar projects be conducted to determine if the Volleyball Rating Scale is meaningful with less skilled female players, and male competitors of various skill levels.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

TABLE OF CONTENTS (CONT.)

TABLE OF CONTENTS (CONT.)

LIST OF TABLES

CHAPTER I

STATEMENT OF THE PROBLEM

I. INTRODUCTION

Scott and French (29) suggest that the application of measurement and evaluation provides a scientific foundation for physical education. The use of measurement in physical education is important in that it is primarily through measurement that the outcomes of teaching or coaching are determined. The physical educator may use test results as a basis for grading, predicting the ability of players, classifying players into teams, comparing players' skill and progress, diagnosing individual weaknesses, and stimulating player interest.

An examination of the methods of measuring ability in sports and athletics indicates three general approaches: (1) the use of standardized skill tests; (2) rating by an expert or panel of experts; (3) subjective evaluation by the teacher or coach. Each of these methods seem to have certain limitations or restrictions when attempting to measure actual playing ability in a game situation.

Standardized skill tests in volleyball are of the non-competitive type, nearly all of which deal with two isolated skills, volleying and serving. The results of skill tests are

often used as an estimate of an individual's ability to perform the skills in the game of volleyball. Considerable research during the past few years has substantiated the theory of task specificity of motor abilities. In general terms, this theory maintains that there is extreme specificity of motor coordination abilities and performances and that isolated physical abilities are specific to that particular task or activity. Bachman (1) suggests that abilities are task specific, both for performance and for motor learning. He suggests that skill tests are not adequate tools for measuring playing ability in a game situation.

Repeated evaluation by experts is generally inconvenient and not realistic although " . . . all measurement in its inception relies on the opinions of experts." (4) It is highly unlikely a physical educator or athletic coach would be able to obtain the services of local experts every time student or player evaluation was desired.

Measurement by the teacher or coach tends to be extremely subjective with personal bias influencing results. Yet this method of measurement in physical education is most frequently employed. Subjective evaluation tends to be masked by the teacher's or coach's personal likes and dislikes, their past experiences, and such factors that consciously or unconsciously enter into this method of evaluation.

How then does one begin to objectively measure and discriminate the playing ability of students or team personnel? More specifically, how does one measure actual playing ability in volleyball objectively and with validity?

II. THE PROBLEM

Statement of the Problem

It is the purpose of this study to develop an instrument that will be valid, objective and practical for discriminating relative volleyball performance in a competitive game situation for skilled female volleyball players.

Sub-Problems

The sub-problems of the study are to determine

1. Which skills are considered of greater importance for successful performance in volleyball and, the relative importance of each.

2. The number of times a rater must use the instrument to obtain high objectivity.

3. The better indicator of playing performance between the Volleyball Rating Scale and the Clifton Single Hit Volley Test.

Importance of the Study

Bovard, Cozens, and Hagman (4) suggest that evaluation is essential for the improvement of teaching techniques and

conditions of learning. Better means of evaluation in athletics and physical education should be constantly sought. Barrow (3) is of the opinion that objective measurement should be used as the primary method of appraisal.

Welch (32:158) states, "Unfortunately, there are just a few volleyball tests which have been subjected to statistical analysis and meet acceptable standards for both validity and reliability." The majority of these tests were devised before 1960. Since then there have been many changes in playing techniques, officiating and rules giving cause to question the validity of some of the present tests. Further objection to these tests is that they generally place little emphasis on the game situation.

The work done on specificity by Henry (18) and others, (1, 10, 22, 26) can be summed up adequately with a quote from Lotter (22:60):

> A test battery can only sample specific abilities and therefore can only be effective in predicting a criterion that involves the specific abilities that are sampled.

The implications of this statement seem to suggest that standardized skill tests are not adequate means of measuring volleyball performance in a game situation. Scott and French (29) suggest skill tests should be as nearly like the game situation as possible. This agrees with the opinion of Barrow (3:26) who states "No test or battery of tests is ever an adequate substitute for the game itself."

There is a great need for a test in volleyball that can discriminate validly, objectively, and practically relative playing performance in a game situation. Presently there are no tests available which measure volleyball performance in a competitive game situation. No test administered during a game situation has been substantiated by scientific evidence to meet the requisites listed by Clarke (8) as criteria for a "good test". Such a tool would prove helpful in objectively revealing team and individual weaknesses and strengths. Such data are needed to indicate those aspects of the game which need further and immediate attention.

Delimitations

This study is not intended to evaluate relative performance of players of average skill in volleyball such as might be encountered in a school situation. Nor is it intended to be used with male competitors. The study will be confined to discrimination of performance of skilled female competitors. The study includes samples from participants at the Second Century Week Volleyball competitions and the Canadian Senior Women's Championships.

Limitations

The devised Volleyball Rating Scale will yield scores that are relative to performance of team players and to the level of competition. Comparison of scores between groups at different

levels of competition will not be meaningful. The development of norms is not an aspect of this study.

No effort will be made to evaluate positional play, desire, sportsmanship, et cetera. Only those skill aspects which can be objectively scored in a quantitative manner will be taken into consideration.

The calibre of opponents cannot be controlled and therefore scores will be dependent upon this factor. Substitutions and amount of time played cannot be controlled during the collection of the data. Therefore, an attempt will be made to control this in the statistical analysis by eliminating those players who do not meet defined requirements of involvement in the game situation.

The assumption is made that reliability of technique is more important than reliability of player performance. Individual performance varies from game to game and even moment to moment depending on such factors as the physiological and psychological state of the performer. However, a problem arises when testing for reliability in that it is difficult to differentiate between player and technique reliability. Should the reliability coefficients not be significant at the .01 level, this will not be considered to weaken the study. Guilford (16:104) states that:

> It is coming to be recognized that validity is much more important than reliability, and in fact, it is possible for a

> test to be sufficiently valid for practical purposes without being very reliable.

In much work in the field of human and animal learning, this is so because fairly gross error attaches to many of the measurements made concerning reliability due to its very nature.

The rank ordering of members of a group results in an ordinal variable. The use of multiple regression equations assumes the data are composed of interval or ratio variables. Thus one of the assumptions underlying the use of Aitken's numerical solution for calculating a multiple regression equation was not met in this study. However, Ferguson (14:14) states that:

> In practice we frequently apply methods appropriate to one class of variable in the statistical analysis of other classes of variables.

He comments further:

> . . . many variables are in fact ordinal, although for statistical purposes they are, quite justifiably, commonly treated as if they were interval or ratio variables. . . . Frequently practical necessity dictates a particular procedure. (14:15)

Definitions of Terms Used

Ace. A service that results in a point immediately or results in a point immediately after initial contact by the opponents.

Error, Violation. Any act which is defined as an error or

violation according to the official rules of volleyball and which is called by game officials.

Good Block. The obstruction of a ball by one or more players who place their hands and/or forearms in the path of the ball being returned by an opponent over the net such that the ball immediately returns to the opponents' court or is deflected into the player's(s') own area in an obvious upward manner.

Good Pass. Any ball played for the first time on a particular side that is directed to another teammate in such a manner that the ball reaches the teammate higher than the shoulders and is between the ten foot back row spiking line and two feet from the net.

Good Set. A ball played with the intent to direct it to a front line attacking player to spike and that is between the ten foot spiking line and two feet from the net; approximately four feet higher than net height; and in the case of a front line setter, between another front line player and the setter. A set can only occur on a second team contact but the second team contact is not necessarily a set.

Good Serve. A service that does not result in an ace nor a poor serve.

Good Spike. Any ball that is met with a one-hand whip-like action (overhead throwing action), taken while in the air, such

that the ball is directed into the opponents' court causing: (1) the opponent to return the ball immediately after one contact (other than a good block); (2) the ball to make contact with the floor immediately in the opponents' court; (3) the opponents not to make the ball playable after initial contact.

Good Return. Any play that results in the ball passing into the opponents' court and that has not been defined elsewhere. (Drive, tip, dump, et cetera.)

Own Area (Block). A team's officially defined court plus an extension of approximately three feet on each sideline.

Playable. Any ball approximately shoulder height off the floor and such that a teammate is within approximately six feet of the ball.

Player. Any competitor who makes more than ten contacts with the ball per game.[1]

Recovery. An emergency play where the ball is played in such a way that it is caused to be returned to the opponents' court or is made playable for a teammate. Had the ball not been recovered, it would have resulted in a point or side-out.

Poor Block. Any ball contacted by intended blockers

---

[1]The first quartile of the total range of contacts per game per subject was used to establish this value. (Q1 = 10.5, range = 1-52, and N = 506).

other than as defined in good block. A poor block is not recorded if the intended blockers make no contact with the ball.

Poor Pass. Any ball played for the first time on a particular side that does not result in a good pass.

Poor Serve. A serve that does not cross the net or that lands out of bounds.

Poor Set. A ball played with the intent to direct it to a front line attacking player for a spike that does not result in a good set. A set can only occur on a second team contact but the second team contact is not necessarily a set.

Poor Spike. Any ball contacted as described in good spike, but that does not result in a good spike.

Poor Return. Any play that does not result in the ball passing into the opponents' court and that has not been defined elsewhere.

CHAPTER II

REVIEW OF THE LITERATURE

I. BACKGROUND

In 1895 William G. Morgan introduced a new indoor game called Mintonette into the physical program of the Young Men's Christian Association center of Holyoke, Massachusetts. Morgan ". . . was searching for an indoor game that was challenging to the young and old alike but not quite so vigorous as the game of basketball" (30:3). Mintonette was soon to become known as volleyball having a large number of participants and spectators throughout the country.

The growth and development of the game was extremely rapid in its early stages with the greatest impetus coming between 1920 and 1930. In 1923 volleyball was adopted as an official activity by the National Amateur Athletic Federation followed in 1928 by the formation of the United States Volleyball Association (U.S.V.B.A.). Further significant happenings added impetus to the game in 1949 when a collegiate unit of competition was held. The first World Championship matches were held in Prague, Czechoslavakia for male competitors. The first women's division was added to the national U.S.V.B.A. champtionships held in Los Angeles, California in the same year.

In 1952 Moscow played host to the second World Championships at which time competition for women was included. In 1955 volleyball was added to the Pan American Games in Mexico City. Spectator interest was extremely high as evidenced by the capacity crowds in attendance.

The International Olympic Committee in 1957 designated volleyball as an official Olympic team sport for men. This was followed in 1962 by the decision to establish Olympic competition for women.

The original game of volleyball has undergone considerable change since its inception when it was generally known as an "old man's" game and a purely recreational sport. Today "power" volleyball is considered to be a "young man's" game destined to become one of the great participant and spectator sports (32).

II. EXISTING VOLLEYBALL TESTS

As early as the 1930's, physical educators were attempting to devise skill tests in volleyball (33). An adequate number of sport technique tests can be found in the literature but there are comparatively few which are valid and objective (4). Skill tests that have been developed have been generally concerned with measurement of two aspects of the game, the serve and the volley.

The lack of objective skill tests which measure playing ability in all aspects of volleyball has caused the teacher and coach to lean heavily towards rating scales and incidence charts for evaluation (30). Subjective evaluation may well accomplish the purpose for which it is intended. However, it would be more valuable if controlled experiments were conducted to determine the validity and objectivity of the evaluation (32).

Criteria for physical education test validation is generally dependent on ratings, or upon objective criteria that are admittedly of low reliability (23). Many of the existing volleyball skill tests that profess to measure playing ability have been validated on the basis of the combined judgment ratings of experts (4).

One of the earliest batteries of tests designed to measure volleyball ability was developed by French and Cooper (15) in 1937. This battery included four tests; the repeated volleys test, net recovery, placement serving, and passing. There was no stated reliability for the battery of tests but the validity was reported as .72 for grade ten to twelve females. Validation was based on four judges' ratings.

In general the remaining skill tests are modifications of the French and Cooper battery. Repeated volley tests are the most common. The Russell and Lange (27) repeated volley test

is reported to have a validity coefficient of .67 based on ratings of seven judges. The score on this test was the number of legally repeated volleys made in thirty seconds against a wall from behind a three foot restraining line. Mohr and Haverstick (25) investigated the reliability and validity of the Russell-Lange test when performed three, five, and seven feet from the wall. The scores were the sum of three trials at each of the distances. The authors reported the highest reliability (.83) was obtained when the test was performed from behind the seven foot restraining line. The best validity (.77) was found by correlating three judges' ratings with the sum of the seven and five foot test scores.

Bassett, Glassow and Locke (33) established a validity coefficient of .51, determined on the basis of the instructor's ratings, for a wall volley test. A six foot restraining line was used at the start of the test only but was ignored after the wall volleying was in progress.

Clifton (9) developed a single hit volley test for women in 1962 that would be consistent with rule revisions. Using forty-five college women, she reported the highest validity coefficient (.70) was obtained when the test was performed from behind a seven foot restraining line and the scores of trials one and two were summed. Again the trials were of

thirty seconds duration. Validity was determined by correlating test scores with rankings of five judges based on one observation of the women in a volleyball game.

Crogen (11) determined the validity of a wall volley test by correlating test results with results obtained from a sixteen team round robin tournament rather than judges ratings. She found the teams made up of players with high test scores won more games than those with low test scores. The validity coefficient was not reported but the reliabilities with one hundred twenty-nine females ranged from .48 to .52 for ten hits and .83 for twenty hits. In this test the time factor was eliminated with the score being the number of consecutive hits executed. No restraining lines were used once the test started.

Liba and Stauff (20) developed a test for the overhead volleyball pass but no attempt was made to establish the validity of the test as a measure of volleyball playing ability.

Miller and Ley (24) have devised incidence charts and rating scales for evaluation of playing performance but their work was not researched to determine validity, objectivity or reliability. Welch (32), Trotter (30), and Emery (13) have suggested methods for evaluating playing performance but these methods are highly subjective and have not been validated

with an external criterion.

Volleyball coaches and instructors have indicated the necessity of game evaluation through the use of statistics but as yet no one has conducted research on methods of gathering statistics to determine validity and objectivity (32).

III. STATISTICAL ANALYSIS AND TEST CONSTRUCTION

Textbooks (8, 28, 29) generally agree that a test can be of maximum effectiveness if it meets the following requirements:

(1) Test validity. The test must measure accurately what it intends to measure. For example, a wall volley test should accurately measure general playing ability.

(2) Test reliability. The test must measure consistently what it intends to measure.

(3) Test objectivity. Two or more individuals, using the same instrument, must obtain similar results.

(4) The results of the tests should be amenable to conversion to normative tables.

(5) The test must be economical of time.

Test Validity

Validity is the degree to which the test measures the quality for which it is to be used. Validity of a test is generally determined by means of descriptive and/or statistical validity.

An established criterion of the elements being measured is selected for comparison with the new test. Clarke (8) suggests that the most frequently used criteria are critical thinking, established tests, subjective ratings and composite criteria factors.

Through logical explanation an investigator can validate a test descriptively by showing that the test does what the descriptive criterion calls for (28). This validity is most frequently termed curricular or face validity. Validating the test statistically involves the use of statistical formulas to correlate the proposed test against a selected test criterion (28). Previously validated skill tests, competitive standings, or judges' ratings are most often used as criteria against which to validate tests statistically.

The proven validity of the test depends upon the degree of relationship between the test and the criterion. If the correlation between the criterion and the test is high (.80 is generally considered significant), they measure the same thing (8). Low validity coefficients between a proposed test and certain criteria may be the result of inaccurate criterion measures and therefore do not prove the test to be invalid.

> For example, certain judgment ratings are known to be inconsistent, and test results compared with such ratings as a criterion, as has been frequently done in the construction of skill tests, would suffer as a result.

> One would naturally, therefore, expect to get lower validity coefficients when such criteria are used (8:28).

Ferguson (14) suggests the Spearman rank-difference method for determining a correlation coefficient rho may be used when two variables are in the ordinal scale and when numbers are small (less than twenty-five).

When no external criteria are available for validating tests, the pooled ordering of judges' rankings may serve as criterion providing there are adequate opportunities for observation and the judges are competent. ". . . all measurement in its inception relies on the opinions of experts" (4) despite the fact that these opinions tend to be inconsistent.

> The objection that ratings are subject to many constant errors is met by proof to the contrary that pooled ratings somehow eliminate the force of these errors and by the fact that certain corrections may be made if necessary (17:280).

Often when no one method of validation seems satisfactory, a combination of several criteria may be used for test validation.

Item Validity

Test validity is influenced by the ability of each test item to discriminate between those who possess skill and those who do not. Efforts should be made to be sure that all items in a test are functioning; that they do something in the way of

measurement. Item analysis procedures enable one to differentiate between the better items and the poorer items.

There are several methods of determining the merits of each part of a test. The Flanagan Index of Discrimination, one of several methods of determining the discriminating ability of test items, yields a correlation coefficient which indicates how well a test item differentiates good and poor performance (28). An item yielding an index of discrimination of twenty or higher is considered to have high discriminating power, providing it meets other criteria, and should be retained in the test (28).

Other methods of determining item validity are functioning of responses, primarily used in knowledge tests, and difficulty rating. These methods are not appropriate for a study of this nature.

The relative importance of each item in a test battery can be determined by the use of multiple regression equation technique (28).

> If the items are all of approximately equal weight (importance), the test author can disregard weighting in setting up the scoring system. In this case the total test score can be computed by converting test item scores to standard scores and then summing. However, if the weightings in the regression equation are unequal, it is best to use the regression equation to compute total performance scores (28:248).

Ferguson (14) describes Aitken's numerical solution for

calculating the required regression weights with more than three variables. This pivotal condensation method and the Doolittle method are most frequently used to obtain information concerning weightings. These statistical procedures indicate which combination and weighting of test items will yield the highest validity with the selected criterion (29).

Reliability

Reliability can be defined as the degree of consistency of results obtained on two or more measurements of the same object or function by the same device and test administrator (8). Reliability of a test will be dependent upon consistent performance of individuals and consistent measurement by the instrument. Reliability coefficients in physical performance tests must be given serious consideration as to their meaning. It is difficult to know if the coefficient indicates consistent individual performance or consistent instrument performance or a combination of both. McCloy (23) suggests that an individual's performance at any one time almost always differs from his average performance over a long period of time and therefore low reliability coefficients may not indicate a weakness in the scientific authenticity of the test. The split-half method of determining reliability might even result in a low reliability coefficient due to varying performance from moment to moment.

Objectivity

Objectivity is dependent upon the ability of two or more examiners to agree when using the same test on identical subjects at the same time (8). Objectivity is determined by correlating the results of the n sets of data obtained by the different investigators.

> . . . tests with high objectivity will also have high reliability . . . . Frequently, therefore, in constructing tests, objectivity only is computed; the assumption is that, if this is satisfactory, reliability is automatically assured (8:31).

CHAPTER III

METHODS AND PROCEDURES

I. THE SAMPLE GROUP

The sample was taken from the female population of competitors in the 1967 Second Century Week Volleyball tournament and the 1967 Canadian Senior Women's Volleyball Championships. A number of competitors were not included in the sample due to the lack of playing time or insufficient involvement in the game. A player was defined as any competitor who made more than ten contacts with the ball per game. It is therefore not possible to obtain results from the entire population. Data were obtained on one hundred sixteen individuals from ten teams in seventeen matches.

The teams competing in the Second Century Week Championships were conference winners from the four womens' university intercollegiate conferences across Canada. The four teams were from the Universities of Manitoba, Toronto, Windsor and New Brunswick. These teams were competing for the Canadian Intercollegiate Athletic Union Championship. Data from seven matches and twenty-five game ratings were obtained at this two day tournament held in Calgary, Alberta, March 6 and 7, 1967. Fifty-two individuals were rated.

The Canadian Senior Womens' Championships were held in

Toronto, Ontario, March 17 and 18, 1967. Nine Canadian regional winners and/or consolation winners competed for the Canadian Volleyball champtionship title. Ratings were obtained on six teams resulting in ten match ratings and twenty-three game ratings. The total number of individuals rated was sixty-four. The teams rated were Marpole I and Marpole II from British Columbia, Winnipeg Buffalos from Manitoba; Toronto Blues, Toronto Plasts, and Ottawa Patro from Ontario. Other teams which competed but not rated were the Cals from Alberta, Hamilton Spartans from Ontario and the Montreal Volleyball Club from Quebec.

A match consisted of the best three of five games at the Second Week Volleyball tournament. A match at the Canadian Senior Womens' Volleyball championships consisted of two games during round-robin play. The final match at this tournament was the best two of three games. A game at both tournaments was won when a team had scored fifteen points and had at least a two point advantage over the opponents.

The order in which teams were rated was determined using random sampling without replacement. After each team had been rated once, random sampling without replacement was again employed for the remaining number of teams. To select teams for rating in semi-final and final games, the teams were randomly selected from those teams involved.

II. THE VOLLEYBALL RATING SCALE

All observations were recorded statistically using the "Volleyball Rating Scale" (See Appendix A). The scale rated eight variables in the game; serve, pass, spike, set, block, recovery, return, and violation/error.

Face validity of the "Volleyball Rating Scale" can be demonstrated by showing that the material covered by the scale is important to successful performance in skilled volleyball competition.

Almost without exception, noted authorities (13, 30, 32) in volleyball agree that the serve is one of the most important plays in the game. DeWeese (32) substantiates this theory with the following reasoning. A team must be able to put the ball legally into play in order to score. That is, a team must be able to serve. When serving, a team can, to some degree, force the opponents to play the ball as the offensive team wishes. By effectively serving the ball, a team has a definite opportunity to put its opponents off balance at least momentarily.

Trotter (30) suggests that the service is potentially the primary offensive weapon in the game of volleyball. "By a team's strengths or weaknesses in the service, a team begins play with an advantage or a disadvantage; hence the serve is

one of the most important offensive plays of the game" (13:35). Each player on a volleyball team should develop serving techniques which will enable them to put the opposing team immediately on the defensive.

It is for these reasons that the "ace" and "good serve" have been included in the Volleyball Rating Scale. The category "poor serve" was included because it was felt that the team that could not put the ball into play could not score points. This was considered to be detrimental to good performance.

Mastery of passing is essential to a good volleyball player as the accurate and well timed pass is very often the key to successful play. Speaking of the passer (referred to as every member of the team) Wilson is quoted by Welch (32:41) as saying:

> He is the man who initiates the attack, and if he does a bad, slovenly job, then the whole pattern of offense is bogged down. It is not a glamorous role, but an essential one.

The greatest variation of teams and players generally occurs in effective execution of the pass. Every player on the team must have volleying skill that enables her to be a good passer. It was, therefore, deemed necessary to include the category of "good pass" and "poor pass" in the Volleyball Rating Scale.

The set or set-up is the second of three contacts that a team has with the ball on its side of the court. The set holds the attack together. Without the effective execution of the set-up, no

volleyball team can muster a winning offense. Clark (32:53) states that "Pass-set-spike should be considered as one continuous play, each equally necessary for a winning offense." Without effective setting a team cannot hope for effective spiking, so essential to potential attacks. The set can logically be defended as an important aspect in the game of volleyball. This fact is further substantiated by the number of skill tests in volleyball centered around the set (15, 30, 33).

Spiking is also a very important skill in the game of volleyball. As Emery (13:32) states, "Master the spike, for it is one of the most effective offensive weapons in volleyball." Walters (32) considers the spike a necessary skill for every member of a team.

"A volleyball team cannot hope to win if they do not block and block well" (32:73). Good volleyball teams must use effective blocks to combat the skill of good spikers. This aspect of the game is probably the difference between the winning and losing teams. Again all players should be capable of doing a good job of blocking the ball (13).

Recovery shots or attempts to regain control of the ball are considered essential skills by Burton. He states (32:80):

> To win in volleyball, you must move defensively so that you may always execute a controlled . . . pass or else learn to use in emergencies a recovery skill which will at least neutralize an opponent's placement advantage or

> a team mate's mistake. It follows that since the former course (perfect performance) is impossible, then the latter is essential to success in competitive play.

Trotter (30) further substantiates this philosophy by suggesting that digs, used as emergency measures, are fast becoming essential ball-handling skills in good volleyball for girls and women.

Errors and fouls are also considered important. In every sport the officials are an important part of the game. Upon their judgment rests the outcome of the game or match. Volleyball is no exception to this observation. Errors and fouls can be extremely costly to a volleyball team. A foot fault during an ace or good serve results in immediate loss of the ball; a foot fault following a good spike negates all previous efforts of that team; a net foul during an attempted block nullifies the blockers efforts; holding, double contacts, redirecting and all other errors result in immediate loss of the ball. And so it is for every error and foul. It is very essential to good volleyball performance that each player minimizes the number of calls made against her due to poor ball handling or fouls. All violations and errors result in loss of the ball or a point for the opponents. This item was therefore negatively weighted in the Volleyball Rating Scale.

Dumps and tips are important for offensive tactics. Although the ultimate aim during the offensive attack is to effectively pass,

set and then spike the ball into the opponents' court, this is not always possible nor desirable. Unless a team has very adept spikers, poorly set balls are better returned with well placed drives to the opponents' court. Very often well executed spikes are neutralized and frequently nullified by the opponent's effective blocking. In such instances "dumps" and "tips" would be far more effective against solid blocking than attempts to spike through or around the block. For these reasons, the category of return has been included in the Volleyball Rating Scale. In effect, this item is an attempt to include all unspecified types of plays across the net.

Team success in any competitive sport demands a mastery of fundamentals. This generality is particularly relevant to volleyball where basic procedures are repeated over and over again, endlessly, on both attack and defense. Fundamentals of volleyball usually listed are these: the serve, the block, and the big three (the games great triumvirate) of the attack: the pass, the set and the spike. Without the effective execution of all these maneuvers no volleyball team can be totally effective.

> It is impossible to say that one aspect is more important or less important than another in obtaining a point or the ball as the case may be (32:43).

The player's number was recorded in the appropriate column or space for spike, pass, set, return, and block. When

the skill was executed as "good" by definition, the player's number was circled. The player's number was not circled when the skill was executed as "poor" by definition.

Recording the three types of serves was done as follows:

(1) A player's number was recorded in the serving box and circled when a serve resulted in an "ace" by definition.

(2) The player's number was recorded in the serving box but not circled when the serve was "good" by definition.

(3) The player's number was recorded around the perimeter of the serving box when the serve was "poor" by definition.

Violations/errors and recoveries were simply recorded by writing the player's number in the appropriate column or space. These two variables cannot be dichotomous. Therefore good and poor recordings were not necessary.

One point was awarded for each variable executed as good by definition while poor executions resulted in no points being awarded. An ace and good serve each received one point. Poor serves were merely recorded as a contact. Violations and errors were recorded as a contact and one point was subtracted.

Item totals for each player were determined and recorded in the appropriate spaces in the upper right hand corner of the rating scale. The number of positive points was summed over all

variables and divided by the total number of contacts made during that game for each player. A game ratio was then determined for each player from which rankings were obtained.

III. EXPERIMENTAL DESIGN

Test Validity

The review of the literature provided little precedence to follow for this method of testing and therefore there seemed to be no single validity criterion which contained all factors involved in the Volleyball Rating Scale. Because of the nature of the Volleyball Rating Scale, the principal method of validation was through the curricular technique although a combination of methods of validation were used.

The first aspect of curricular validity was demonstrated by showing that the material covered by the Volleyball Rating Scale is important to successful performance in skilled volleyball competitions. It was also shown that the Volleyball Rating Scale closely simulates a real game situation.

Spearman's coefficient of rank correlation rho (14) was used to correlate the test results of each rated match with the pooled rankings of a panel of judges. All rho coefficients were transformed to Fisher's z, summed, divided by the number of correlations, and transformed back to a rho coefficient.

Panel of Judges

The panels of judges were selected on the basis of the following criteria:

(1) Previous coaching experience within the past five years.

(2) Playing experience in skilled volleyball competition within the past five years.

(3) Experience as a judge or panel member for a player selection committee.

(4) Currently a nationally qualified official.

Only persons meeting at least one of these requirements served on the judges' panel

Two female physical education professors from the University of Calgary served as judges for the Second Century Week competitions. Both judges had coached female intercollegiate volleyball teams within the past five years. One was coaching the women's team at the University of Calgary. One of the judges played for the Calgary Cals Volleyball team which had represented Alberta in the Canadian Volleyball competitions for the past three years. This same judge had also been selected to the Canadian Women's All Star volleyball team two of the past three years and represented Canada in the 1967 Pan American Games Volleyball Competitions.

The panel at the Canadian Senior Womens Volleyball

competitions consisted of two female judges. Each of the judges had participated in previous Canadian competitions within the past five years.

The panels of judges met before competitions began to discuss and jointly decide upon those skills of the game they considered essential for good performance in volleyball. Following this they independently observed the players of a randomly selected team for an entire match. At the conclusion of the match they ranked the players on the basis of their overall performance and contribution to the team effort for that match. Two major assumptions were made concerning the judges. First, that the judges' rankings were valid, and second, that the judges' rankings were reliable.

Item Validity

Each item's discriminating power was determined by the Flanagan technique (29). An item was considered to discriminate if it yielded a Flanagan index of .20 or better as suggested by Scott (28, 29). Item analysis was made by comparing the item point ratios with the total Volleyball Rating Scale ratios of the top third players and the bottom third players.

The relative importance of each variable in the Volleyball Rating Scale was determined by using Aitken's method of pivotal condensation (14) to develop a multiple regression equation even

though the variables were ordinal and not interval or ratio. The criterion was the judges' average rankings.

Reliability

The reliability of the Volleyball Rating Scale as a measure of relative volleyball performance was determined by correlating the Volleyball Rating Scale rankings of game one with those of game two, using the Spearman rank correlation coefficient rho. Low reliability coefficients were not considered to weaken the instrument if validity had been shown.

> It is coming to be recognized that validity is much more important than reliability, and, in fact, it is possible for a test to be sufficiently valid for practical purposes without being very reliable (16:104).

Objectivity

Objectivity was determined by correlating the Volleyball Rating Scale scores obtained by the author with the rating scale scores obtained by two independent raters at the Second Century Week tournament. The independent raters were students attending the University of Calgary. One rater had played volleyball for the University of Calgary for two years. The other rater had been manager of the university team for one year.

The Spearman rho correlation technique was used to determine an objectivity coefficient. The rho coefficients were converted to Fisher's z scores, a mean z determined and reverted back to a

rho coefficient for comparison of the total objectivity ratings.

The number of times a rater must use the Volleyball Rating Scale to obtain accurate results was determined by the number of games an observer had to rate to obtain an objectivity coefficient significant at the .01 level. Spearman's coefficient of rank correlation was used.

Comparison of a Skill Test, the Volleyball Rating Scale and the Judges' Evaluations

Validity of the Clifton Single Hit Volley Test as a means for discriminating relative performance of skilled players in a competitive game was determined by correlating the results of the test with the rankings obtained by the judges. The Clifton skill test was administered to the four women's teams competing in the Second Century Week competitions.

A test for significant differences between the validity coefficients of the Clifton Single Hit Volley Test and the Volleyball Rating Scale, as obtained by correlating with the judges' rankings, was made to determine which test is more valid as an indicator of relative playing performance of skilled female volleyball players.

All formulas used in the study may be found in Appendix C.

CHAPTER IV

RESULTS AND DISCUSSION

I. THE VOLLEYBALL RATING SCALE

The Volleyball Rating Scale, which includes the fundamentals of volleyball plus returns, and errors/violations, is considered to have curricular validity. As a measure of volleyball performance, all overt acts that could be objectively scored in a quantitative form in an actual game situation were used in the Volleyball Rating Scale. The only uncontrolled situations affecting the rankings obtained by the Volleyball Rating Scale were the calibre of opponents and the amount of playing time of each team member. It is therefore proposed that the Volleyball Rating Scale embodies a realistic situation.

Table I presents a mean game comparison of successful contacts to total contacts for each variable included in the Volleyball Rating Scale. The Second Century Week Tournament teams averaged a total of eighty-nine (89) contacts per game, forty-nine (49) of which were considered to be successful contacts. The teams evaluated at the Canadian Senior Womens' Volleyball Championships averaged a total of one hundred thirty-six (136) contacts per game. Seventy-four (74) of the one hundred thirty-six (136) contacts were considered good. The mean number of contacts for each variable was greater at the Canadian Senior Womens'

Championships than at the Second Century Week Tournament.

Table II indicates that the mean Canadian Senior Womens' Volleyball Championships team ratio was superior to the mean Second Century Week Tournament team ratio in the pass, set and block. The mean Second Century Week Tournament team ratio was superior to the mean Canadian Senior Womens' Volleyball Championships team ratio in total game, spike and return.

Table III compares the Second Century Week Tournament final team standings with the mean game ratio ranking obtained by the Volleyball Rating Scale for each variable as well as the overall mean game ranking. Comparisons are also shown for the teams that competed in the Canadian Senior Womens' Volleyball Championships.

The University of Toronto won the Second Century Week Tournament. This team was also considered by the Volleyball Rating Scale to be the better team in the tournament in terms of mean game performance. However, the University of Toronto only ranked first on the Volleyball Rating Scale variables set, ace, good serve and recovery. The University of Manitoba placed second at the Second Century Week Tournament and ranked second in mean game performance on the Volleyball Rating Scale. The University of Manitoba ranked first on the

TABLE I

MEAN GAME COMPARISONS OF SUCCESSFUL TO TOTAL CONTACTS

	Total		Spike		Pass		Set		Recov-	Return		Block		Viola-	Ace	Good Serve	Poor Serve
	T	S	T	S	T	S	T	S		T	S	T	S				
University of Manitoba	99	59	7	5	17	9	14	9	.5	6	4	5	2	1.5	4	13	2
University of Toronto	116	71	16	6	32	15	24	19	1	13	9	5	1	2	5	14	2
University of Windsor	62	27	4	2	26	5	8	5	1	6	5	5	1	3	1	9	1
University of New Brunswick	82	42	8	3	13	9	22	10	.4	10	8	3	2	4	2	10	1
SECOND CENTURY WEEK TOURNAMENT MEAN	89	49	9	4	22	9	17	10	.7	9	6	4	1	2	3	11	1
Marpole I	150	80	24	9	43	19	29	23	.5	12	7	14	7	4	3	15	3
Marpole II	140	70	18	7	44	16	31	24	3	12	7	4	4	5	2	16	2
Toronto Blues	189	112	26	15	56	25	37	27	.7	16	10	16	10	4	3	23	4
Winnipeg Buffaloes	144	85	26	14	44	23	31	23	.5	12	10	3	1	.7	3	36	4
Toronto Plasts	88	47	10	2	30	12	19	13	1	4	1	11	8	.5	2	7	2
Ottawa	103	51	11	4	39	15	18	14	0	12	9	2	2	4	1	9	5
CANADIAN SENIOR WOMEN'S VOLLEYBALL CHAMPIONSHIPS MEAN	136	74	19	8	43	18	27	20	1	12	7	8	5	3	3	18	3

TABLE I

TABLE II

MEAN GAME RATIOS OF SUCCESSFUL TO TOTAL CONTACTS

Team	Total	Spike	Pass	Set	Return	Block
University of Manitoba	.58	.71	.52	.64	.66	.40
University of Toronto	.61	.37	.46	.79	.69	.20
University of Windsor	.43	.50	.19	.62	.83	.20
University of New Brunswick	.51	.37	.69	.45	.80	.66
SECOND CENTURY WEEK TOURNAMENT	.55	.44	.40	.58	.66	.25
Marpole I	.53	.37	.44	.79	.58	.50
Marpole II	.50	.38	.36	.77	.68	1.00
Toronto Blues	.59	.56	.44	.73	.91	.91
Winnipeg Buffaloes	.59	.53	.52	.74	.83	.33
Toronto Plasts	.53	.20	.40	.68	.25	.72
Ottawa	.49	.36	.38	.77	.75	1.00
CANADIAN SENIOR WOMEN'S VOLLEYBALL CHAMPION-SHIPS	.54	.42	.41	.77	.58	.62

TABLE III

COMPARISON OF TOURNAMENT RANKINGS TO MEAN ITEM AND GAME RANKINGS

Team	Tournament Ranking	Volleyball Rating Scale	Spike	Pass	Set	Return	Block	Violation	Ace	Good Serve	Poor Serve	Recovery
University of Toronto	1	1	3.5	3	1	3	3.5	2	1	1	3.5	1.5
University of Manitoba	2	2	1	2	2	4	2	1	2	2	3.5	3
University of New Brunswick	3	3	3.5	1	4	2	1	4	3	3	1.5	4
University of Windsor	4	4	2	4	3	1	3.5	3	4	4	1.5	1.5
Marpole I	1	3.5	4	2.5	1	5	5	4	2	4	3	4.5
Toronto Blues	2	1.5	1	2.5	5	1	3	4	2	2	4.5	2
Marpole II	3	5	3	6	2.5	4	1.5	6	4.5	3	1.5	3
Winnipeg Buffaloes	4	1.5	2	1	4	2	6	2	2	1	4.5	4.5
Toronto Plasts	5	3.5	6	4	6	6	4	1	4.5	6	1.5	1
Ottawa	6	6	5	5	2.5	3	1.5	4	6	5	6	6

spike and violations and second on pass, set, block, ace and good serve.

The University of New Brunswick was rated third on mean game performance by the Volleyball Rating Scale. This team finished third in the Second Century Week Tournament. The University of New Brunswick ranked first on the pass, block and poor serve; second on the return; and third on the spike, ace, and good serve. The University of Windsor placed fourth at the Second Century Week Tournament. The Volleyball Rating Scale also ranked this team fourth on mean game performance.

There was no significant correlation ($r = .46$) between team placement at the Canadian Senior Womens' Volleyball Championships and the mean game ranking obtained by the Volleyball Rating Scale. The Marpole I team finished first at the tournament but only ranked first on the Volleyball Rating Scale in the set. The Toronto Blues, who placed second at the Canadian Senior Womens' Volleyball Championships, were ranked first by the Volleyball Rating Scale on spike and return and tied for first on mean game performance.

II. STATISTICAL VALIDITY

For each match rated, the judges' average ranking was correlated in two different ways with the rankings obtained by the

Volleyball Rating Scale. In the first instance a Spearman's coeffficient of rank correlation (rho) was determined by ranking the players on the overall match ratio without concern for the number of games played in the match. The second rho was determined by assigning ranks with concern for the number of games each person played in the match. Those persons who had played more than half the total number of games per match were ranked according to the overall match ratio following which those persons who had played less than half the number of games per match were ranked. A summary of the validity coefficients as determined by each method is shown in Table IV. Individual coefficients were converted to Fisher's z transformation, added, averaged and converted back to a Spearman rho.

Table IV indicates that the average rho (method 1) for all rated matches was .109 as opposed to the average rho of .470 as determined by method 2. These results were determined from seventeen validity coefficients with a mean N of 7. Both coefficients are well below values required for significance at either the .05 level or .01 level. The average rho for all Second Century Week matches, as determined by method 1, was .004 compared to the average rho of .590 as determined by method 2. Again neither coefficient is significant at either the .05 or .01 level. These

TABLE IV

VALIDITY COEFFICIENTS AS DETERMINED

BY TWO SEPARATE METHODS

Team	Mean Number of Individuals	Number of matches	Rho #1	p	Rho #2	p
All teams	7	17	.109	.8>p>.9	.470	.2>p>.3
All Second Century Week teams	8	7	.004	p<.9	.590	.05>p>.1
All Canadian Senior Women's Volleyball Championships teams	6	10	.181	.7>p>.8	.375	.4>p>.5
'University of Manitoba	8	2	.058	p<.9	.725*	.01>p>.05
'University of Toronto	7	2	-.259	p<.9	.470	.2>p>.3
'University of New Brunswick	8	2	.016	p<.9	.695*	.01>p>.05
'University of Windsor	7	1	.392	.3>p>.4	.584	.1>p>.2
-Marpole I	7	2	.297	.5>p>.6	.650	.1>p>.2
-Marpole II	7	2	.551	.1>p>.3	.645	.1>p>.2
-Toronto Blues	7	2	.020	p<.9	.420	.2>p>.3
-Buffaloes	6	2	-.195	p<.9	-.195	p<.9
-Plasts	6	1	.130	.8>p>.9	.130	.8>p>.9
-Ottawa	6	1	.200	.7>p>.8	.200	.7>p>.8

' Second Century Week teams
- Canadian Senior Women's Championship teams
* Significant at the .05 level

Rho #1 No concern for number of games played by each player per match.
Rho #2 Concern for number of games played by each player per match.

rhos were established from seven validity coefficients with a mean N of 8.

The Canadian Senior Women's Championship matches produced an average rho #1 of .181 and an average rho #2 of .375 both being well short of significance at either the .05 or .01 level. These rhos were determined from ten validity coefficients with a mean N of 6.

Although the statistical analysis has indicated that the Volleyball Rating Scale is not a valid instrument with which to discriminate relative volleyball playing performance, regardless of the method used to obtain rho, it is suggested that the lack of proven validity with these panels of judges does not weaken the curricular validity of the scale. A coefficient of concordance W, used to determine the degree of agreement among judges, could not be established due to the fact that only two persons composed each of the panels of judges. The low validity coefficients may be partially attributed to a lack of agreement between the judges.

Table IV would also seem to suggest certain weaknesses were evident in the average rankings of the judges. The wide divergence between the rhos, as determined by the two separate methods, seems to indicate the judges tended to rank favorably those persons whom they saw most frequently. Individuals who played less than half the games per match tended to be ranked

lower regardless of their playing ability. This observation seems to be more apparent in the matches of the Second Century Week tournament where a match consisted of a minimum of three games and a maximum of five games. In the Canadian Senior Women's Championships a match consisted of only two games except in the final where a match was the best of two out of three games.

It is of interest to note that a higher average rho #2 was obtained from the Second Century Week tournament than from the Canadian Senior Women's Championships. This may possibly be due to the fact that the panel of judges at the first tournament had a better opportunity to evaluate performance due to the greater number of games per match. This perhaps suggests a panel of judges needs a longer period of time to observe performance before arriving at rankings which are meaningful in terms of the Volleyball Rating Scale.

The larger coefficients obtained by method #2 seems to indicate that the judges were inclined to rank players according to a coaches' player selection rather than actual performance. That is, players who participated in the greatest number of games were ranked higher than those players who played a minimum number of games.

The validity coefficients may have been insignificant due

to the relatively equal weighting of the items on the Volleyball Rating Scale. As Welch (32) points out it is difficult to say which aspect(s) of the game is more important in obtaining a point or the ball and therefore all items were either given a positive or negative value of one. All rhos were calculated using a constant weight of absolute one for each variable.

It was felt that the judges probably considered some items more important to team success than others. Therefore the relative importance of each variable in the Volleyball Rating Scale was determined for both tournaments using the judges' match rankings as the criterion. As outlined in Ferguson (14), Aitken's method of pivotal condensation was used to establish a multiple regression equation which would produce a system of weights providing the best estimate of the criterion. Data from the Second Century Week produced seven correlation coefficients for match rankings on each variable with the judges' match rankings as well as seven intercorrelation coefficients for each variable. The coefficients in each category were transformed to Fisher's z, summed, averaged and converted back to a Spearman rho. The procedure was the same for the data from the Canadian Senior Women's Championships.

The multiple correlation coefficient (R) from the Second

Century Week data was .656 as compared to an R of .832 obtained from the Canadian Senior Women's Championships. Both R's were significant at the .01 critical level.

Table V presents the determined regression coefficients for each variable as well as the average correlation coefficient of each variable with the judges' match rankings from both tournaments. Rounded weight values for each variable are also shown.

Table V clearly shows that the two panels of judges did not consider all items equally important for successful performance. These results also indicate that the two panels were not in agreement concerning the relative importance of each item. One panel considered the order of importance for successful performance to be good serve, ace, pass, set, spike, block, poor serve and return. The second panel ordered the items for successful performance as good serve, pass, spike, ace, return, poor serve, set and block.

The two separate panels also weighted the items very differently although the range of rounded weights was comparable. The range of weights for panel #1 was -2 to +4 and for panel #2 was -4 to +8. Each panel considered a good serve to be most important to successful performance. Whereas the first panel considered the ace to be second in terms of relative importance

TABLE V

MULTIPLE REGRESSION COEFFICIENTS

Item	SECOND CENTURY WEEK			CANADIAN SENIOR WOMEN'S CHAMPIONSHIPS		
	Rho with judges	Regression coefficient	Rounded weight	Rho with judges	Regression coefficient	Rounded weight
Spike	.18	.063	1	.16	.247	3
Pass	.30	.160	2	.30	.686	7
Set	.21	.091	1	-.05	-.434	-4
Return	.07	-.151	-2	.07	.090	1
Block	.48	.003	1	.13	-.386	-4
Ace	.49	.254	3	.51	.196	2
Good Serve	.65	.382	4	.45	.805	8
Poor Serve	-.05	(-).237	(-)2	.17	(-).037	(-)1

and gave it a weight of three, the second panel gave the item a weight of two and considered it fourth in terms of importance. Passing was considered to be second most important to successful performance by the second panel of judges and received a weighting of seven, only one point less than a good serve. The first panel, on the other hand, weighted the item with two, half as much as a good serve, and considered passing to be third most important.

The most striking differences, both in terms of relative importance and weighting, occurred in setting and blocking. One panel weighted the two items negative four, the other positive one. The negative regression coefficients do not imply that either panel of judges necessarily considered some skills detrimental to performance. The negative weights do suggest that some of the predictors were acting as suppressor variables. The use of the regression coefficients are only meaningful when used as a unit. If any predictors (skills) were eliminated from the Volleyball Rating Scale, a new set of regression coefficients must be determined in order for them to be meaningful.

If panels of judges such as used in this study are considered to be valid external criteria against which to validate a new testing instrument, then further use of the Volleyball Rating Scale would necessitate multiplying successful or

unsuccessful item attempts (as the case may be) by the appropriate weight, summing separately the numerators and denominators over all items and then dividing the numerator by the denominator to obtain a game or match ratio. However, when two panels of judges, such as used in this study, produce such divergent item weights, it tends to make judges suspect for validating new testing instruments.

As will be noted in Table V, page 47, a positive regression coefficient was obtained for poor serve. This resulted because of the inverse ordering on this item. In actual fact this item bears a negative weighting on both scales.

As hypothesized, the panels of judges did consider some skills more important than others and therefore the low validity coefficients obtained by correlating the Volleyball Rating Scale's match rankings with those of the judges are not considered to weaken the value or validity of the Volleyball Rating Scale.

It was desired to know if the validity coefficients obtained by correlating the Volleyball Rating Scale rankings with those of the judges would be improved if the rounded weights were used to determine the rankings on the Volleyball Rating Scale. A new rho was therefore determined for each rated match using the regression coefficients. Match rhos were converted to Fisher's z, summed, averaged and reverted back to a rho. Table VI presents these results.

TABLE VI

DETERMINED VALIDITY COEFFICIENTS

USING REGRESSION COEFFICIENTS

Team	Mean Number of Individuals	Number of matches	Rho #1	p	Rho #2	p
All teams	7	17	.085	p<.9	.420	.2>p>.3
All Second Century Week teams	8	7	.230	.5>p>.6	.690*	.01>p>.05
All Canadian Senior Women's Volleyball Championship teams	6	10	-.017	p<.9	.165	.7>p>.8
'University of Manitoba	8	2	.110	.8>p>.9	.785*	.01>p>.05
'University of Toronto	7	2	.435	.3>p>.4	.675	.05>p>.1
'University of New Brunswick	8	2	.175	.6>p>.7	.590	.1>p>.2
'University of Windsor	7	1	.130	.7>p>.8	.670	.05>p>.1
+Marpole I	7	2	-.081	p<.9	-.081	p<.9
+Marpole II	7	2	-.045	p<.9	-.145	p<.9
+Toronto Blues	7	2	-.740	p<.9	.015	p<.9
+Buffaloes	6	2	.315	.5>p>.6	.315	.5>p>.6
+Plasts	6	1	.930*	.01>p>.05	.930*	.01>p>.05
+Ottawa	6	1	-.31	p<.9	-.31	p<.9

' Second Century Week teams

+ Canadian Senior Women's Championship teams

* Significant at the .05 level

Rho #1 No concern for number of games played by each player per match.

Rho #2 Concern for number of games played by each player per match.

The new mean validity coefficient for all rated teams was found to be .085 when no consideration was given to the number of games each player played per match. The validity coefficient was .420 when consideration was given to the number of games each person played per match. A comparison of Table IV (page 42) and Table VI indicates that mean rho #1 and #2 for all teams was higher when the rounded weights were not used to determine match rankings.

The mean rho #2 obtained for all teams competing at the Second Century Week tournament was .690 when the rounded weight factor was used to determine match rankings on the Volleyball Rating Scale. This coefficient was slightly higher than the rho #2 of .590 when all items were weighted absolute one. The mean rho #2 obtained for all teams competing at the Canadian Senior Women's Volleyball Championships was .165 when the rounded weight factor was used to determine match rankings on the Volleyball Rating Scale. This rho #2 was considerably lower than the rho #2 obtained when the Volleyball Rating Scale items were weighted equally.

Rho #2 was higher for the University of Windsor, the University of Manitoba, and the University of Toronto when the Volleyball Rating Scale items were weighted. However, rho #2

was significant for the University of New Brunswick when the Volleyball Rating Scale items were weighted equally. The use of relative item weights resulted in insignificant rho #2's for this team.

Rho #1 was higher for three teams at the Second Century Week Tournament when items were weighted according to the regression coefficients. Only rho #1 for the University of Windsor did not improve when the items were weighted.

Rho #1 and rho #2 were reduced for four teams competing at the Canadian Senior Women's Volleyball Championships when the items were weighted according to the regression coefficients. Rho #1 and rho #2 were improved for Toronto Plasts and Winnipeg Buffaloes when the items were weighted.

The rounded weight factors produced from the data obtained at the Second Century Week tournament appears to be a better weighting system than that obtained for the Canadian Senior Women's Volleyball Championships when the rho is determined after considering the number of games played per match by each player.

III. ITEM ANALYSIS

The Flanagan Index of Discrimination as outlined by Scott (28) was used to determine the ability of each Volleyball Rating Scale item to discriminate between good and poor performers. Table VII indicates the index of discrimination obtained for each

TABLE VII

ITEM DISCRIMINATION

Item	Flanagan Index	Significance
Set	.35	significant
Pass	.39	significant
Return	.32	significant
Block	.27	significant
Violations/errors	.46	significant
Ace	.36	significant
Poor serve	.32	significant
Good serve	.13	not significant
Spike	.18	not significant

of the variables. An item was considered to discriminate if it yielded a Flanagan Index of .20 or better.

All items except spiking and good serving yielded significant indexes of discrimination. One possible explanation for the low discriminating ability of good serving may lie in the necessity of each team member being able to legally put the ball into play. Without the basic skill of serving, a person would not likely be selected for the team. A second possible explanation is that by definition a good serve tends to include a wide variety of serves and therefore might not be expected to discriminate between good and poor performers. There does not seem to be any logical reason why spiking ability should not discriminate relative performance. Possibly the definition of a good spike is too stringent. Either the spiking ability of the players is too weak for the definition or the defense is too strong for the definition. The index of discrimination for the spike was close to the .20 Flanagan Index which is required for significance.

Theoretically both the spike and good serve items should be eliminated from the Volleyball Rating Scale if the Flanagan index of discrimination is the sole basis for inclusion of items. However, both the curricular validity and the multiple regression equation suggest that each of these items is important for good

volleyball performance and therefore it is proposed that these two items remain in the Volleyball Rating Scale.

IV. RELIABILITY

A reliability coefficient was determined by correlating the ranking of game one with those of game two as obtained from the Volleyball Rating Scale using the Spearman rank correlation coefficient rho. Only matches in which the same personnel played both game one and game two were used to determine this coefficient. Reliability coefficients from eleven matches were transformed to Fisher's z, summed, averaged and reverted to a rho. The resultant reliability coefficient was .395 with a mean N of 6, considerably short of significance at the critical levels of .01 or .05.

Because curricular validity has been shown for the Volleyball Rating Scale, the low reliability coefficient obtained was not considered to weaken the instrument. As Ferguson states (14:288), " . . . low reliability does not necessarily invalidate a technique as a device for valid inferences." Guilford (16:104) supports this stand when he comments,

> It is coming to be recognized that validity is much more important than reliability, and, in fact, it is possible for a test to be sufficiently valid for practical purposes without being very reliable.

Clarke (8) has defined reliability as the degree of consistency of results obtained on two or more measurements of the same

object or function by the same device and test administrator. The key words in this definition are " . . . two or more measurements of the same object or function." For the Volleyball Rating Scale to be a reliable instrument, it must consistently rank each playing member of a team in the same order from one game or match to the next. Theoretically, this means that should a team member have an exceptionally good performance and rank number one, she must rank number one every time thereafter.

Adequate reliability coefficients have been reported for standardized skill tests which measure a series of non-competitive isolated drills. However, the Volleyball Rating Scale attempts to determine relative performance in a game situation. Individual performance can be expected and does vary in competition from moment to moment, game to game and match to match depending on a vast number of variables such as fatique, warm-up, motivation, and opponents. As McCloy (23) suggests, an individual's performance at any time almost always differs from his average performance over a long period of time. It is therefore suggested that a test for reliability may be reporting the degree of consistency in individual performance from one game to the next rather than consistency in instrument performance. The low reliability coefficient obtained on the Volleyball Rating Scale may not indicate a weakness in the scientific authenticity of the Volleyball Rating Scale.

V. OBJECTIVITY

Objectivity was determined by correlating the Volleyball Rating Scale match scores for the Second Century Week tournament with the Volleyball Rating Scale match scores obtained by two independent raters. Table VIII indicates the obtained coefficients for match play between each of the independent raters and the author as well as the averaged objectivity coefficients.

As can be seen from Table VIII the average objectivity coefficient in match ratings for all teams was significant beyond the .01 level. This same statement is true for each of the independent raters. With the exception of the University of New Brunswick rho, all coefficients were significant at or beyond the .01 level. The level of significance of the University of New Brunswick match was .05.

The number of times a rater must use the Volleyball Rating Scale before obtaining objectivity at the .01 level for a match was determined. These results are given in Table IX.

Table IX indicates that both raters were able to obtain objectivity coefficients significant at the .01 level during their first use of the Volleyball Rating Scale. It will be noted that the third and fifth rated matches did not obtain objectivity at the .01 level but rather the .05 level. Two factors that might

TABLE VIII

OBJECTIVITY COEFFICIENTS FOR MATCH PLAY

Team	N	Rater #1	Rater #2	Average
University of Toronto	8	.900*	.835*	.870*
University of Manitoba	8	.950*	.930*	.940*
University of New Brunswick	8	.780**	.650**	.722**
University of Windsor	9	.836*	.780*	.810*
All teams	8	.900*	.850*	.876*

* Significant at the .01 level
** Significant at the .05 level

TABLE IX

SEQUENTIAL OBJECTIVITY COEFFICIENTS

Match Order	N	Team Contacts per match	Rater #1	Rater #2	Average
1	9	186	.836*	.780*	.810*
2	8	432	.881*	.923*	.905*
3	8	210	.780**	.650**	.725**
4	8	195	.980*	.930*	.962*
5	8	257	.822**	.804**	.813**
6	7	474	.943*	.860**	.913*

* Significant at the .01 level
** Significant at the .05 level

influence the level of significance obtained for any particular match would be the type of match being played and the experience of the recorders for the raters. The first objectivity coefficient was obtained from a match in which the maximum number of contacts per player in any one game was fifteen as opposed to some matches which produced maximum contacts of fifty-two per game for a particular player. Obviously the fewer the total number of contacts per game the longer the rater has to make a decision concerning the contact and the more time the recorder has for marking the scale.

Although the objectivity analysis is primarily to determine if independent raters interpret the definitions of the Volleyball Rating Scale in a similar manner as the author, the experience of the recorders might influence the level of significance obtained for the objectivity coefficients. Each rater had a recorder for each match but the recorders were not always the same person. If the recorder was unfamiliar with the recording sheet or was unable to keep pace with the rater in a well played match, then the objectivity coefficients may be low due to the recorder rather than the rater. There was no way to check out this hypothesis during this study.

It was desired to determine at what stage in a match the

independent raters were able to obtain objectivity at the .01 level. Table X is an analysis of averaged game objectivity coefficients. As indicated in Table X Rater #2 was able to obtain an objectivity coefficient significant at the .01 level during the first game of a match while it took Rater #1 two games to obtain a significant objectivity coefficient.

VI. COMPARISON OF THREE METHODS OF EVALUATION

The Clifton Single Hit Volley Test is a standardized skill test designed to measure overall ability in volleyball by means of evaluating a player's ability to repeatedly volley a ball against a wall from behind a seven foot restraining line. This skill test was administered to all teams competing in the Second Century Week tournament. The results of the skill test were compared to the assigned rankings of the judges to determine the validity of the skill test to discriminate relative playing performance of skilled players in a competitive game situation. Spearman rhos were determined for each match, converted to Fisher's z, summed, averaged and reverted to a rho. The average rho was .165 as determined from six matches with a mean N of 7. This coefficient is not significant at either the .01 or .05 level.

According to this panel of judges, the Clifton Single Hit Volley Test is not a valid tool for measuring volleyball ability

TABLE X

GAME OBJECTIVITY COEFFICIENTS

Game Number	N	Rater #1	Rater #2
1	6	.910**	.948*
2	7	.907*	

** Significant at the .05 level
* Significant at the .01 level

in a competitive game situation. This seems to suggest that neither the Volleyball Rating Scale nor the skill test would be valid means for determining relative performance. The only other method for evaluating performance seems to be subjective evaluation if the validity coefficients obtained by correlating with the rankings of the panels of judges used in this study are meaningful.

It was desired to know whether there was a significant difference between the rho of .165 of the skill test judges and the rho of .470 of the Volleyball Rating Scale judges. A test for significant differences resulted in a t of .96 which indicated the two rhos are not significantly different. According to this panel of judges one method of evaluating relative game performance is as good or as poor as the other.

Assuming the Volleyball Rating Scale to be a valid instrument with which to measure relative game performance, as determined from curricular validity, a correlation for validity was determined between the Volleyball Rating Scale and the skill test. The validity coefficient was found to be .132. This is not significant at the .01 or .05 level. If the face validity of the Volleyball Rating Scale is accepted, then the Clifton Single Hit Volley Test is not a valid means of discriminating relative playing performance in volleyball.

VII. DISCUSSION

Three methods presently exist for validating a new evaluative instrument. In this study, the most significant method of validation was through curricular validity. It was shown that the Volleyball Rating Scale contained items considered by authorities to be important to successful volleyball performance in a game situation. The scale also entailed a realistic game situation which helped to substantiate the face validity of the scale.

A second means for validating new measuring techniques is to obtain significant correlation between the instrument and some external criterion. Panels of experts and standardized skill tests are most frequently used as the external criteria in physical education. Each of these methods was used to obtain a validity coefficient for the Volleyball Rating Scale. In both instances insignificant validity coefficients were obtained which, in the opinion of some people, would tend to discredit the total validity of the Volleyball Rating Scale. However, a closer look into the use of panels of judges and standardized skill tests for validating new testing instruments may suggest why such low coefficients were obtained for the Volleyball Rating Scale.

It is the opinion of this writer that the validity coefficients of .109 and .470, obtained in this study by correlating the results

of the Volleyball Rating Scale with the rankings of a panel of two judges, do not weaken the face validity of the instrument. It is suggested that a panel of two is not sufficient for obtaining reliable and valid rankings. When only two persons serve on the panel, there is no way to determine the extent of agreement between the judges and only the averaged rankings can be used as the external criterion. Unless the rankings of each of the judges are in close harmony, the average of the two has very little meaning. As Clarke (8) has indicated, judgment ratings are known to be inconsistent. When these judgment ratings are used as the criterion for validating a new instrument, low coefficients obtained may be the result of inaccurate criterion measures rather than any inherent weaknesses in the proposed instrument.

Had the panel of judges consisted of three to five persons a statistical test for the amount of agreement among the judges could have been performed. A resultant significant degree of agreement would make it easier to place more faith in any obtained validity coefficients. It is suggested that the obtained validity coefficients between the Volleyball Rating Scale scores and the panel of judges were insignificant because there were insufficient numbers of judges on the panels. It seems probable that the validity of the Volleyball Rating Scale would be somewhat higher than the presently achieved values if the panels were composed of larger numbers of judges.

There was some evidence to suggest that the panels of judges were not consistent in their interpretation of the relative importance of the skills of volleyball. Determined multiple regression equations indicated that the two panels held considerably different views on the weighting and relative importance of the items in the Volleyball Rating Scale. If these experts cannot agree and cannot be consistent in their opinions then it would be difficult to attain validity coefficients which are significant and meaningful.

The obtained correlation coefficients of the scale items with the separate panels of judges also suggests that perhaps personal bias, conscious or unconscious, plays an important part in how the judges rank relative performance. The judges composing the first panel had each played a considerable amount of skilled volleyball, primarily as setters. The second panel consisted of one person who had played strictly as a setter and a second who had played strictly as a spiker. It will be noted in Table V, page 47, that the set correlated .21 with the first panel of judges but -.05 with the second panel. This incongruity may be the result of the judges' personal bias.

Interpretation of the obtained validity coefficients of the present study should be made after consideration is given to the fact that many standardized skill tests, whose validity have been

determined from judges' ratings, are centered around the serve and the volley. The Volleyball Rating Scale category of good serve was not considered to discriminste good and poor performances according to the judges used in this study. It seems peculiar that some judges would consider serving ability to discriminate between good and poor performers in volleyball while other judges would not. There is obviously some inconsistency in this observation.

As a standardized skill test, the Clifton Single Hit Volley Test purports to measure volleyball playing ability. The test is reported as having a validity coefficient of .70 as determined by correlating the test scores with the rankings of five judges on one observation of the sample of a volleyball game. The rho of .165 obtained between the results of the skill test in this study and the match rankings of the two judges was insignificant at the .05 level. Assuming the skill test to be reliable and valid, this result seems to suggest that the panel of judges used in this study was not capable of accurately evaluating relative performance. Or conversely, assuming the judges to be valid criteria, the skill test is not an adequate means for evaluating relative performance in a game situation. In either case, the use of two judges as external criteria for validating new measuring instruments appears to be a dubious practice. If validating new tests by means of judges or

skill tests seems to be a weak practice, the only other method is curricular validity.

As discussed earlier, the determined reliability coefficient of .395 for the Volleyball Rating Scale, insignificant at the .05 level, was not considered to weaken the scientific authenticity of the rating scale. When using player rankings from one game to the next as the basis for comparison, it is difficult to determine whether or not the coefficient represents the reliability of player performance, the instrument or a combination of both. The use of films would make the task of interpreting the reliability coefficient somewhat easier. If films were taken of several matches in tournament play, these could be shown at a later date and new ratings obtained at this time. The ratings obtained from the films could then be correlated with the ratings obtained during the actual game or match. In this way there is no doubt that the players' performance would be exactly the same from one test period to the next and that any obtained coefficient would in fact indicate the reliability of the instrument.

The results of the present study provide a new instrument which contains face validity, is objective, and economical in terms of time and cost, for the evaluation of relative performance of skilled female volleyball competitors. Results obtained on player

performance in a game situation using the Volleyball Rating Scale would seem to be more meaningful than results obtained from standardized skill tests for the same purpose. The research of Triplet, Berridge and Gordon as reported by Clarke (10) would imply that volleyball performance in a series of non-competitive drills would be quite different from performance in a competitive, integrated game situation.

Present use of the Volleyball Rating Scale is relatively restrictive in that results would only be meaningful if they were obtained on skilled female players. However, coaches of intercollegiate or senior women's teams should find the instrument useful for any of the following purposes.

1. The coach can use the Volleyball Rating Scale as an aid in selecting team personnel. Well controlled scrimmages would allow for the application of the Volleyball Rating Scale. The results would provide an estimate of relative performance and thereby assist a coach in selecting team players.

2. A coach could use the Volleyball Rating Scale for diagnostic purposes in terms of individual or team strengths and weaknesses during the competitive season. Each item on the Volleyball Rating Scale can be analyzed separately for the team or for each individual. The analysis would assist the

coach in developing future practice plans by indicating those items requiring more or less practice. If the team ratio in passing is low during a particular match or tournament, much time should be spent in future practices attempting to improve this particular skill.

3. Tabulation of match results obtained from the Volleyball Rating Scale could also serve to motivate team members. A person who consistently scores relatively low in a particular item should be motivated by the results to improve her personal performance in that item.

It seems possible that further study with the Volleyball Rating Scale might broaden the extent to which the Volleyball Rating Scale could be used, perhaps with little or no modifications. Investigation should be conducted to determine whether the instrument is valid, objective and practical when used with less skilled competitors, and with male players. A physical education teacher might then be able to use the results of the scale as a means for evaluating relative pupil performance in a volleyball game.

CHAPTER V

SUMMARY AND CONCLUSIONS

I. SUMMARY

The principal objective of this study was to develop a valid, objective and practical measuring instrument that would discriminate relative volleyball performance of skilled female players in a competitive game situation. Secondary purposes of the investigation were to determine:

1. Whether panels of judges considered all items on the rating scale to be equally important for successful volleyball performance and, if not, the relative importance of each;

2. The number of times a rater must use the instrument to obtain objectivity at the .01 level of significance;

3. Which of the Volleyball Rating Scale and the Clifton Single Hit Volley Test is the better indicator of relative game performance.

Data were collected from two separate volleyball tournaments. Fifty-two individuals were rated during seven matches, consisting of twenty-five games, at the Second Century Week tournament. Another sixty-four ratings were obtained in ten matches of twenty-three games at the Canadian Senior Women's Volleyball Championships. All observations were recorded using the Volleyball Rating Scale.

Face validity of the instrument was shown by demonstrating that the items included in the Volleyball Rating Scale were important to successful performance in skilled volleyball competitions. Curricular validity was further substantiated by showing that the Volleyball Rating Scale embodies a game situation.

Statistical validity did not substantiate the demonstrated curricular validity of the Volleyball Rating Scale when results of the Volleyball Rating Scale were correlated with the averaged rankings of a panel of judges using Spearman's coefficient of rank correlation rho. The low statistical validity was suggested to be due to unequal weighting of items by the panels of judges. Whereas the Volleyball Rating Scale considered each item to be equally important to successful performance, the panels of judges placed greater emphasis on some aspects of the game than others.

The discriminating power of nine scale items was determined by the Flanagan technique. Seven items were found to discriminate.

Reliability of the Volleyball Rating Scale rankings was reported to be low when correlating first and second game rankings of all matches in which the same players participated in both games.

The Volleyball Rating Scale was proven to be an objective

instrument as determined by correlating the rankings of the scale with those obtained by two independent raters. The number of matches and games in which a rater must use the Volleyball Rating Scale to obtain objective results at the .01 level of significance was determined using Spearman's coefficient of rank correlation rho.

The match rankings of the judges were correlated with the rankings obtained from the Clifton Single Hit Volley Test to determine a validity coefficient for the skill test. In this study, the obtained coefficient, which was insignificant, suggested that the skill test was not a valid means of evaluating relative playing performance of skilled players or conversely, that two judges do not constitute an acceptable panel of experts.

A test for significant differences between the obtained validity coefficients of the Clifton Single Hit Volley Test and the Volleyball Rating Scale was made to determine if one method of evaluation provided more accurate results than the other or if the two were comparable in terms of evaluating relative playing performance of highly skilled female players.

The validity coefficient was found to be .109 using the average rankings of the panel of judges and the match rankings obtained by the Volleyball Rating Scale without concern for the number of games played by each player in the match. A second

validity coefficient of .470 was obtained by correlating the average rankings of the judges with the match rankings of the Volleyball Rating Scale when consideration was given to the number of games each person played in a match. Both coefficients were statistically insignificant.

The use of multiple regression techniques indicated that the two separate panels did not consider the items on the Volleyball Rating Scale to be of equal importance. The panel of judges at the Second Century Week tournament ranked and weighted the items in terms of importance as follows: good serve (4), ace (3), pass (2), set (1), spike (1), block (1), return (-2), and poor serve (-2). The panel at the Canadian Senior Women's Championships weighted and ranked the items as good serve (8), pass (7), spike (3), ace (2), return (1), poor serve (-1), block (-4), and set (-4).

It was found that set, pass, return, block, violations, ace and poor serve discriminated between good and poor performances. Good serve and spike did not significantly discriminate performance.

The reliability of the Volleyball Rating Scale, determined by correlating the rankings of game one with those of game two using the Spearman rank correlation coefficient rho, was found to be .395 which was considerably short of significance at the .05 level.

Objectivity coefficients between the author's Volleyball Rating Scale scores and those obtained by two independent raters were .900 and .850 with the average being .876. Both these objectivity coefficients were significant beyond the .01 level of confidence. Each rater obtained objectivity after rating one match. One rater was able to obtain objectivity after the first game of a match while the other required two games to attain significance at the .01 level.

Validity of the Clifton Single Hit Volley Test as a means of discriminating relative performance in a competitive game was determined by correlating rankings obtained by the judges. The determined coefficient was .165 which was not significant at the .05 or .01 level. A test for significant differences between the determined validity coefficients of the skill test-judges and Volleyball Rating Scale-judges was found to be insignificant ($p < .05$).

On the basis of the face validity of the Volleyball Rating Scale, a correlation for validity was determined between the Volleyball Rating Scale and the Clifton Single Hit Volley Test. The skill test was found to be inadequate as a means of discriminating relative performance of skilled female volleyball players in this study ($p < .05$).

II. CONCLUSIONS

Within the limitations of the statistical procedures employed,

the experimental design, the samples investigated, and the personnel serving as judges, conclusions that may be stated from the results of this study are as follows:

1. The Volleyball Rating Scale was demonstrated to possess face validity. The Volleyball Rating Scale was not a valid instrument in the opinion of the panels of judges for evaluating relative playing performance of skilled female participants.

2. The objectivity of the Volleyball Rating Scale was found to be significant. Objectivity coefficients significant at the .01 confidence interval indicated that independent raters were able to effectively use the Volleyball Rating Scale without extensive training or experience in volleyball.

3. Two separate panels of judges did not consider the Volleyball Rating Scale items to be of equal importance. A wide divergence was found to exist in the relative importance accorded each of the Volleyball Rating Scale items by the two panels.

4. Evaluation of relative playing performance by means of the Volleyball Rating Scale was found to be practical in that only two persons, one rater and one recorder, were required to obtain objective results.

5. No significant differences were found between the Clifton Single Hit Volley Test and the Volleyball Rating Scale as methods of determining relative game performance.

6. The Volleyball Rating Scale can be used either separately or in conjunction with other evaluation methods for determination of relative performance of skilled female volleyball players.

7. The Volleyball Rating Scale may be used for determination of individual or team strengths and weaknesses with respect to the items recorded on the scale.

8. The Volleyball Rating Scale has two major advantages over conventional methods of volleyball evaluation. First, it measures a realistic competitive situation and, second, it is diagnostic in that the source of errors is clearly demonstrated.

III. RECOMMENDATIONS

The previous discussions and conclusions from the results of this study have led the writer to make the following recommendations:

1. That further investigations be made with the instrument using panels of judges consisting of at least five members to determine if statistical validity can be improved.

2. That further study be undertaken to determine if the relative importance and weighting of the Volleyball Rating Scale items can be accurately established.

3. That similar projects be conducted to determine if the Volleyball Rating Scale is meaningful with (a) less skilled

female players and (b) male competitors of various skill levels.

4. That the reliability of the instrument be further investigated. The use of films would be extremely effective for this purpose in that player performance could be held constant.

5. That further study is necessary to determine why the spike and good serve items did not adequately discriminate between good and poor performances.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Bachman, John C. "Specificity vs. Generality in Learning and Performing Two Large Muscle Motor Tasks," Research Quarterly, 32: 3-11, March, 1961

2. Barrow, H.M. "Basketball Skill Test," The Physical Educator, 16: 26-27, March, 1959.

3. _____. "The What and How of Testing," The Physical Educator, 12: 8-9, March, 1955.

4. Bovard, J.F., Cozons, F.W., and Hagman, E.P., Tests and Measurements in Physical Education. Philadelphia: W.B. Saunders Company, 1949.

5. Brady, George F. "Preliminary Investigations of Volleyball Playing Ability," Research Quarterly, 16: 14-17, March, 1945.

6. Broer, Marion. "Reliability of Certain Skill Tests for Junior High School Girls," Research Quarterly, 29: 139-145, May, 1958.

7. Butler, Willie Mae. "Comparison of Two Methods of Measuring the Degree of Skill in the Underarm Volleyball Serve," Research Quarterly, 32: 261-262, May, 1961.

8. Clarke, H. Harrison. Application of Measurement to Health and Physical Education. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1967.

9. Clifton, Marguerite A. "Single Hit Volley Test for Women's Volleyball," Research Quarterly, 33: 208-211, May, 1962.

10. Cratty, Bryant J. Movement Behavior and Motor Learning. Philadelphia: Lea and Febiger, 1964.

11. Crogen, Corinne. "A Simple Volleyball Classification Test for High School Girls," The Physical Educator, 4: 34-37, 1943.

12. Edwards, Allen L. Experimental Design in Psychological Research. Toronto: Holt, Rinehart and Winston, 1964.

13. Emery, Curtis Ray. Modern Volleyball. New York: The Mac-Millan Company, 1956.

14. Ferguson, George A. Statistical Analysis in Psychology and Education. Toronto: McGraw-Hill Company, Inc., 1966.

15. French, E. and Cooper, B. "Achievement Tests in Volley-ball for High School Girls," Research Quarterly, 8: 150-157, May, 1937.

16. Guilford, J.P. Fundamental Statistics in Psychology and Education. Toronto: McGraw-Hill Company, Inc., 1965.

17. _____. Psychometric Methods. New York: McGraw-Hill Company, Inc., 1936.

18. Henry, Franklin, and Smith, Loon E. "Simultaneous vs. Separate Bilateral Muscular Contractions in Relation to Neural Overflow Theory and Neuromotor Specificity," Research Quarterly, 32: 42-46, March, 1961.

19. Kay, H. Kenner. "A Statistical Analysis of the Profile Technique for the Evaluation of Competitive Basketball Performance," Unpublished Master's thesis, University of Alberta, Edmonton, 1966.

20. Liba, Marie R., and Stauff, Marilyn R. "A Test for the Volleyball Pass," Research Quarterly, 34: 56-63, March, 1963.

21. Lindeburg, Franklin A. "A Study of the Degree of Transfer Between Quickening Exercises and Other Coordinated Movements," Research Quarterly, 20: 180-195, May, 1949.

22. Lotter, Willard. "Specificity or Generality of Speed of Systematically Related Movements," Research Quarterly, 32: 55-62, March, 1961.

23. McCloy, C.H. "A Statistical Device for the More Accurate Validation of Tests," The Physical Educator, 2: 62-64, December, 1941.

24. Miller, Donna Mae, and Ley, Katherine L. Individual and Team Sports for Women. New York: Prentice-Hall, Inc., 1955.

25. Mohr, D.R., and M.J. Haverstick. "Repeated Volleys Test for Women's Volleyball," Research Quarterly, 26: 179-184, May, 1955.

26. Nelson, D.O. "Studies of Transfer of Learning in Gross Motor Skills," Research Quarterly, 28: 364-374, December, 1957.

27. Russel, N. and E. Lange. "Achievement Tests in Volleyball for Junior High School Girls," Research Quarterly, 11: 33-41, December, 1940.

28. Scott, M. Gladys (ed.). Research Methods in Health Physical Education Recreation. Washington, D.C.: American Association for Health, Physical Education and Recreation, 1959.

29. _____, and Esther French. Measurement and Evaluation in Physical Education. Dubuque, Iowa: Wm. C. Brown Company Publishers, 1959.

30. Trotter, Betty Jane. Volleyball for Girls and Women. New York: The Ronald Press Company, 1965.

31. Ulrich, Joseph. "An Investigation of Existing Methods of Rating Pupils in Physical Education," The Physical Educator, 3: 34-40, October, 1942.

32. Welch, Edmund J. (ed.). How to Play and Teach Volleyball. New York: Association Press, 1960.

33. West, Charlotte. "Wall Volley Skill Test," D.G.W.S. Volleyball Guide,1963-65, 33-37.

APPENDICES

APPENDIX A

VOLLEYBALL RATING SCALE

VOLLEYBALL RATING SCALE

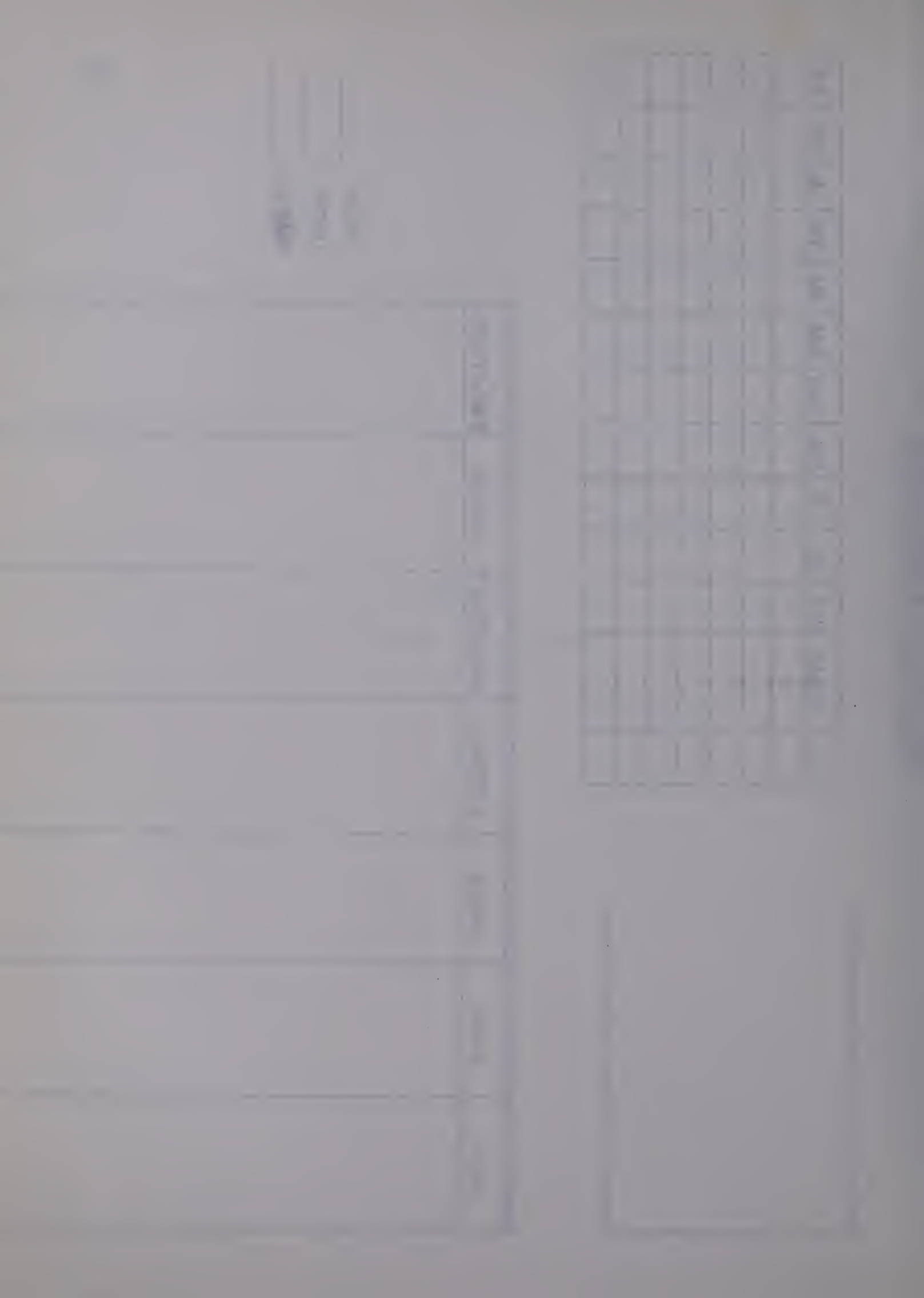
#	NAME	TR	Sp	P	Set	Rec	Ret	Bl	Vio	A	GS	PS

SET	PASS	SPIKE	RETURN	RECOVERY	BLOCK	VIOLATION

TEAM _____

DATE _____

RATER _____



APPENDIX B

MULTIPLE REGRESSION MATRIXES

TABLE XI

MULTIPLE REGRESSION MATRIX

SECOND CENTURY WEEK

	Judge	Spike	Pass	Set	Return	Block	Ace	Good Serve	Poor Serve
Judge	1.00	.18	.30	.21	.07	.48	.49	.65	-.05
Spike	.18	1.00	.04	.00	-.15	.26	.08	.07	.16
Pass	.30	.04	1.00	-.06	.33	.28	.01	.18	.12
Set	.21	.00	-.06	1.00	.02	.13	.19	.20	-.04
Return	.07	-.15	.33	-.02	1.00	-.10	.23	.05	.15
Block	.48	.26	.28	.13	-.10	1.00	.32	.51	.37
Ace	.49	.08	.01	.19	.23	.32	1.00	.73	-.15
Good Serve	.65	.07	.18	.20	.05	.51	.73	1.00	-.09
Poor Serve	-.05	.16	.12	-.04	.15	.37	-.15	-.09	1.00

TABLE XII

MULTIPLE REGRESSION MATRIX

CANADIAN FINALS

	Judge	Spike	Pass	Set	Return	Block	Ace	Good Serve	Poor Serve
Judge	1.00	.16	.30	-.05	.07	.13	.51	.45	.17
Spike	.16	1.00	.12	.33	.13	.40	.03	.11	.01
Pass	.30	.12	1.00	.12	-.08	.41	.18	-.39	-.11
Set	-.05	.33	.33	1.00	.09	.13	.13	.31	-.54
Return	.07	.13	-.08	.09	1.00	.18	.19	.13	.35
Block	.13	.40	.41	.13	.18	1.00	.44	.06	.13
Ace	.51	.03	.18	.13	.19	.44	1.00	.46	-.06
Good Serve	.45	.11	-.39	.31	.13	.06	.46	1.00	.04
Poor Serve	.17	.01	-.11	-.54	.35	.13	.06	.04	1.00

APPENDIX C

FORMULAE USED

FORMULAE USED

1. Spearman's Coefficient of Rank Correlation Rho.

Rho is a measure of association which requires that both variables be measured in at least the ordinal scale so that the objects or individuals under study may be ranked in two ordered series (16). Ferguson (14) presents the rho formula as follows:

$$\rho = 1 - \frac{6\,d^2}{N\,(N^2 - 1)}$$

Where:

N = the number of subjects or objects studied.

d^2 = the sum of the square of the differences between paired ranks.

2. Significance of the Difference Between Two Correlation Coefficients for Correlated Samples.

Ferguson (14) presents the above formula as follows:

$$t = \frac{(r_{12} - r_{13})\sqrt{(N - 3)\,(1 - r_{23})}}{\sqrt{2(1 - r_{12}^2 - r_{13}^2 - r_{23}^2 - 2r_{12}r_{13}r_{23})}}$$

Where:

N = the number of subjects or objects studied.

APPENDIX D

DESIGN

TABLE XIII

DESIGN

	Volleyball Rating Scale	Judges	Skill Test	Author's Ratings	Independent Ratings
Test Validity (Spearman rho)	/////	/////			
Skill Test Validity (Spearman rho)		/////	/////		
Test for Significant Differences	/////		/////		
Item Discrimination (Flanagan (.20))	U=match rank L=upper 1/3 lower 1/3				
Objectivity (Spearman rho)				/////	/////
Game Objectivity (Spearman rho)				/////	/////
Item Weighting (Multiple Regression Equation)	/////	/////			
Test Reliability (Spearman rho)	Game #1 vs Game #2				

APPENDIX E

DESIGN AND SAMPLE DESCRIPTION

TABLE XIV

DESIGN AND SAMPLE DESCRIPTION

	Second Century Week Games (S.C.W.) Calgary, March 6-7	Canadian Senior Women's Championships (C.S.W.C.) Toronto, March 17-18	Total
Teams	Four (4) Intercollegiate Conference Championship Teams Competing for Interconference Championship.	Junior, Intercollegiate or Senior Women's Teams (9 Regional Winners) Competing for Canadian Championship.	Thirteen (13) Teams.
Validity Ratings	Seven (7) match ratings. Twenty-five (25) game ratings. Fifty-two (52) individuals rated.	Ten (10) match ratings. Twenty-three (23) game ratings. Sixty-four (64) individuals rated.	Seventeen matches. Forty-eight games. One hundred sixteen individuals.
Judges	Panel of two (2) judges evaluating each determined match.	Panel of two (2) judges evaluating each determined match.	Two separate panels of two (2).
Objectivity	Two (2) independent raters evaluating six (6) matches.	None.	
Skill Test	Four teams administered the Clifton Single Hit Wall Volley Test.	None.	